用 Pearson-Ⅲ概率分布推算重现期年最大日雨量

林两位 王莉萍

(福建省漳州市气象局,漳州 363000)

摘要 许多工程的设计需依据给定重现期的降水极值,概率推算是其中一种重要的方法。Pearson-Ⅲ型曲线常用于拟合降水频数分布,该文简要介绍了用 Pearson-Ⅲ型概率分布推算重现期年最大日雨量的基本方法。采用漳州市 10 个测站 1961 ~ 2003 年最大日雨量资料,计算给定重现期的最大日雨量。运用常见的办公软件 Excel 2000 的函数计算功能,构造 Pearson-Ⅲ概率分布函数公式,实现软件的快捷计算。初步结果为:在全市 10 个站的拟合中,有 7 个站拟合效果较好,3 个站拟合效果较差。文中就提高 Pearson-Ⅲ概率分布计算精度方面进行了讨论。

关键词 Pearson-Ⅲ概率分布 重现期 年最大日雨量 Excel

引言

在各类自然灾害造成的总损失中,气象灾害引起的损失占 70 %以上。暴雨所引发的自然灾害是较常见的,对于一些重现期很长,如百年一遇的特大暴雨,虽然发生的几率小,但若出现则可能造成毁灭性的灾害,因此许多工程的设计需依据给定重现期的降水极值。气候极值都是相对于某一统计时段而言,如某要素的极大值为 X_p,若大于或等于 X_p 的事件平均每 N 年出现一次,则 X_p 就称为 N 年一遇的极大值,其重现期就是 N 年。在水文如水库和洪水设计中的可能最大降水,有些采用物理成因分析方法,但许多降水极值仍通过概率分析的方法来获得。目前推算可能最大降水量的方法较多[1],基本上可归纳为两种:第一种为水文气象法,称为可能最大降水,简称 PMP;第二种为频率计算法,也称数理统计法,计算一定重现期的年最大日雨量。如陈建昌[2]用 Jenkinson 法计算了山东省及地区 16 个台站年最大日雨量,13 个台站拟合较佳,Jenkinson 法对于特大年最大日雨量一般拟合较差;翟裕宗[3]用三江平原 25 个台站的降水资料,利用耿贝尔公式计算 20 年一遇的年最大日雨量与实际对照,有 6 个台站拟合较差;马开玉[4]用南京市 57 年最大日降水量资料,试用 Pearson-Ⅲ型曲线拟合降水频数分布,经计算 Pearson 拟合分歧度 χ^2 值很小,说明拟合南京年最大日降水量频数分布很好。

有关研究和实践证实,Pearson-Ⅲ型概率分布曲线能较好拟合许多地区的暴雨频数分布,本文用该方法计算漳州市年最大日雨量分布,并在 Excel 2000 电子表格上构造 Pearson-Ⅲ概率分布函数公式,实现快捷计算。

1 Pearson-Ⅲ概率分布

Pearson-Ⅲ分布(以下简称 P-Ⅲ分布)具有广泛的概括和模拟能力,在气象上常用来拟合年、月的最大风速和最大日降水量等极值分布[5]。其概率密度函数和保证率分布函数分别为:

$$f(x) = \frac{\beta^{\alpha}}{\Gamma(\alpha)}(x - x_0)^{\alpha-1}e^{-\beta(x-x_0)}$$

$$\alpha > 0, \quad x \geqslant x_0 \tag{1}$$

$$P(x \geqslant x_p) = \frac{\beta^{\alpha}}{\Gamma(\alpha)}\int_{x_p}^{\infty}(x - x_0)^{\alpha-1}e^{-\beta(x-x_0)}\,dx \tag{2}$$

其中参数 x_0 为随机变量 x 所能取的最小值,α 称为形状参数,β 为尺度参数,$\Gamma(\alpha)$ 是 α 的伽玛函数。用矩法可得 3 个参数的表达式:

$$\alpha = 4/c_s^2 \tag{3}$$

$$\beta = 2/\sigma c_s \tag{4}$$

作者简介:林两位,男,1970 年生,在读研究生,高级工程师,从事天气预报及气候应用研究,Email:llwok2002@yahoo.com.cn
收稿日期:2004 年 6 月 21 日;定稿日期:2004 年 8 月 25 日

$$x_0 = m(1 - \frac{2c_v}{c_s}) \tag{5}$$

式中 m 为数学期望,σ为均方差,c_s 为偏态系数,c_v 为变差系数。其估计量分别为:

$$\hat{m} = \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \tag{6}$$

$$\hat{\sigma} = s = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2} \tag{7}$$

$$\hat{c}_v = \hat{\sigma}/\hat{m} = s/\bar{x} \tag{8}$$

$$\hat{c}_s = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^3 \Big/ \left|\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2\right|^{3/2} \tag{9}$$

据式(2)、(3)、(4)、(5)可知,理论保证率 P 与 x_p 的函数关系由 3 个参数 m、c_v、c_s 所决定,令

$$\Phi = \frac{X - m}{\sigma} = \frac{X - m}{mc_v} \tag{10}$$

称之为 P-Ⅲ分布的离均系数,记 φ为 Φ的取值,则有

$$x = \varphi\sigma + m = m(\varphi c_v + 1) \tag{11}$$

$$P(\Phi \geqslant \varphi_p) = \frac{\left|\frac{2}{c_s}\right|^{4/c_s^2}}{\Gamma(\frac{4}{c_s^2})}\int_{\varphi_p}^{\infty}(\varphi + \frac{2}{c_s})^{\frac{4}{c_s^2}-1} e^{-\frac{2}{c_s}(\varphi+\frac{2}{c_s})} d\varphi$$

式中 $P(\Phi \geqslant \varphi_p)$ 仅与参数 c_s 有关,给定后,即确定了 P 与 φ_p 的关系。通常情况下,是根据估计值$\hat{c}_s$,查数理统计表得 P 与 φ_p 的对应值,再用式(11)计算 P 与 x_p 的对应值。因偏态系数$\hat{c}_s$ 含有 3 阶样本矩,易造成较大的抽样误差,样本实测值 c_s 与分布真值c_s 之间可能会有较大差异,故常需对拟合的线型进行验证并对估计参数值$\hat{c}_v$、$\hat{c}_s$ 作适当调整,以取得较理想的分布曲线,称为适线法。对 P-Ⅲ分布的检验可使用 χ^2 拟合优度检验。

2　在 Excel 上计算 Pearson- Ⅲ概率分布

Excel 具有强大的数据分析功能和丰富的数据表达能力[6],Excel 2000 在数理统计方面的能力很强,只要充分利用其函数功能,通过构造函数,即可完成多数统计功能。

因 Excel 中无 P- Ⅲ分布的概率密度函数及分布函数,要实现在“电子表格”中计算,必须构造相应的数学表达式。经分析,Excel 2000 现有的统计函数中,Γ分布与 P- Ⅲ分布关系最密切,可以用来构造所需的函数。

Γ分布的概率密度和分布函数为:

$$f(t) = \frac{1}{\mu^v\Gamma(v)} t^{v-1} e^{-t/\mu}$$
$$x \geqslant 0,\ \mu > 0,\ \alpha > 0 \tag{12}$$

$$P(t \geqslant 0) = \frac{1}{\mu^v\Gamma(v)}\int_0^{\infty} t^{v-1} e^{-t/\mu} dt \tag{13}$$

对比分析 P- Ⅲ分布和 Γ 分布的概率密度和分布函数,令式(12)中的 $v = \alpha$, $\mu = 1/\beta$, $t = x - x_0$,则式(12)转换为式(1),同理,可将式(13)转化为式(2),须注意的是,Γ分布函数求的是左尾概率,也就是累积概率,而 P- Ⅲ分布函数是右尾概率,因此在变量替换后,还要用 1 减去左尾概率转为右尾概率。

在 Excel 2000 中,GAMMADIST 函数可以计算 Γ 分布的概率和累积概率。

语法:GAMMADIST(x,α,β,c)

x 为用来计算 Γ 分布的数值;α、β为分布参数,如果 β=1,函数 GAMMADIST 返回标准 Γ 分布;c 为一逻辑值,决定函数的形式,如果为 TRUE,函数返回累积分布函数,如为 FALSE,则返回概率密度函数。

另外,GAMMAINV 函数为伽玛分布的累积函数的逆函数,如果 P = GAMMADIST(x,...),则 GAMMAINV(P,...) = x。

语法:GAMMAINV(P,α,β)

P 为伽玛分布的概率值,α、β为分布参数。如果 β=1,函数 GAMMAINV 返回标准伽玛分布。

利用上述 Γ 分布的两个函数,通过变量替换,可构造出 P- Ⅲ分布的概率 P=1 - GAMMADIST($x - x_0$,α,1/β,TRUE),可用于知道 x 值求其出现概率 P 的情况;而其反函数为 x = GAMMAINV(1 - P,α,1/β) + x_0,用于求某一概率 P 相对应的 x 值。

3　实例计算

应用上述方法,容易求出某站的气象极值的 P- Ⅲ分布。以漳州市共 10 个台站的年最大日雨量分布为例简要介绍。气象记录从 1961 至 2003 年,样本数共有 43 个,将每年的最大日雨量挑出,顺序输入到 Excel 2000 工作表中的不同单元格,如有王岩的 Q3X 地面观测查询软件,用项目整编方法及数据导出功能,可快速实现将历史的年最大日雨量数据导出到 Excel 工作表中。芗城区的 43 个样本为 109.6、125.7、215.9、98.1、104.7、90.0、103.8、101.1、91.9、90.9、96.2、181.3、88.0、125.2、97.4、81.3、50.5、136.9、98.2、112.8、100.5、77.2、110.3、

95.9、195.4、124.8、71.2、80.5、91.4、108.0、105.4、132.2、118.2、72.3、126.5、127.6、209.9、194.0、90.1、181.6、119.7、97.4、109.5,最大值为 215.9,最小值为 50.5。

3.1 确定 Pearson- Ⅲ概率分布的 3 个参数

用 AVERAGE 函数计算出样本的均值 $m = 114.86$,用 STDEVP 函数算出均方差 $\sigma = 37.42$,用式(8)可得 $c_v = 0.33$,SKEW 为 Excel 中计算分布的偏斜度的公式,与式(9)略有不同,若 43 年的样本为 $r_1, r_2, ..., r_{43}$,则 $c_s = \text{SKEW}(r_1, r_2, ..., r_{43})(n-1)(n-2)/n^2$,其中 $n = 43$,由此算出 $c_s = 1.21$。P- Ⅲ分布的 m、c_v 和 c_s 3 个参数算出后,分布曲线即已确定。

3.2 计算给定重现期的年最大日雨量

P- Ⅲ分布的 m、c_v 和 c_s 确定后,利用式(3)、(4)及(5)计算得 $\alpha = 2.72$, $\beta = 0.044$, $x_0 = 53.17$,用 $x = \text{GAMMAINV}(1 - P, \alpha, 1/\beta) + x_0$,输入各种概率 P,即可算出对应 P- Ⅲ分布的年最大日雨量的理论值 x。如计算百年一遇的年最大日雨量,在概率 P 单元格输入 0.01,即得 233.01(mm)。

若求某一年最大日雨量 x 值所对应的概率 P,则用公式 $P = 1 - \text{GAMMADIST}(x - x_0, \alpha, 1/\beta, \text{TRUE})$,在 x 单元格输入 295.7,得 P 值为 0.001,即为千年一遇。

同理可算出其余 9 个台站的分布情况。表 1 列出漳州市 10 个站的几种不同重现期的年最大日雨量的理论值。

表 1 漳州市 10 个站不同重现期年最大日雨量理论值

台站	m/mm	c_v	c_s	50 年 mm	100 年 mm	200 年 mm	500 年 mm	1000 年 mm
芗城	114.9	0.33	1.21	213.3	233.0	252.3	277.2	295.7
龙海	114.3	0.38	1.02	224.2	245.1	265.4	291.5	310.9
东山	122.8	0.46	1.26	274.1	304.7	334.6	373.5	402.5
诏安	137.6	0.38	0.76	264.1	286.4	307.8	335.0	355.0
云霄	150.8	0.35	0.70	275.8	297.4	318.0	344.2	363.4
平和	122.7	0.32	0.65	217.6	233.7	249.2	268.7	283.0
南靖	123.8	0.36	1.80	250.7	279.7	308.6	346.6	375.2
华安	101.0	0.29	0.95	175.0	188.8	202.1	219.3	231.9
长泰	103.0	0.30	1.20	183.3	199.3	215.0	235.2	250.2
漳浦	140.1	0.42	0.85	287.0	313.6	339.3	372.1	396.2

注:m 为年最大日雨量平均值。

在计算中,诏安、云霄、平和及漳浦的 x_0 均为负数,而降水量不可能出现负值,须有 $x_0 \geqslant 0$,根据式(5)调整 c_v 和 c_s,使 $c_s = 2c_v$,即让 $x_0 = 0$ 再进行计算。

3.3 Pearson- Ⅲ概率分布拟合检验

算出 P- Ⅲ分布后,把各(P, x_p)点绘在概率格纸上,连接成一光滑曲线,即为 P- Ⅲ理论保证率曲线,再用经验保证率曲线对比验证,其方法为,对实测 n 年降水极值,按从大到小的顺序排列 $x_1 \geqslant \dots \geqslant x_m \geqslant \dots x_n$,经验保证率分布函数:

$$P_n(x) = P(X \geqslant x) = \begin{cases} 0 & x > x_1 \\ \dfrac{m}{n+1} & x_m \geqslant x > x_{m+1} \\ \dfrac{n}{n+1} & x \leqslant x_n \end{cases}$$

将样本$(P(x_m), x)$也一一点绘在概率格纸上,绘制为平滑曲线,与理论保证率曲线进行比较,目测其拟合程度。若拟合不佳,则需进行调整。目测拟合情况人为因素较大,年最大日雨量的实际总体分布是否符合 P- Ⅲ分布,最好使用客观的分布拟合检验。以下采用 χ^2 拟合优度检验:

①根据年最大日雨量样本的频次分布情况分成 k 个区间段 (a_i, a_{i+1}), $i = 1, 2, ..., k$,区间段的划法应根据具体情况来确定,要求每个区间段内的频数不少于 5,否则进行区间合并,合并后为 $k = 5$,即分为 5 段;

②计算概率

$$P_i = F(a_{i+1}) - F(a_i) \qquad i = 1, 2, \cdots, k$$

样本量 n 与 P_i 的乘积 Z_i 为理论频数。

③计算样本观测值 $x_1, x_2, ..., x_n$ 落在区间段 $(a_i, a_{i+1}]$ 中的个数 Z_{0i}, $i = 1, 2, ..., k$, Z_{0i} 称为观测频数。

④按下式计算 χ^2 值

$$\chi^2 = \sum_{i=1}^{k} \frac{(Z_{0i} - Z_i)^2}{Z_i}$$

⑤P- Ⅲ分布有 3 个未知参数,自由度为 5 - 3 - 1 = 1。

⑥给定显著性水平 α,按特定的自由度确定临界值 $\chi^2_\alpha(1)$。

⑦进行判断,如果 $\chi^2 < \chi^2_\alpha(1)$,则认为样本服从 P- Ⅲ分布。

从漳州市 10 个站的 χ^2 值(表 2)可看出,全市 10 个站中诏安、云霄、平和、南靖、华安、长泰和漳浦等 7 个站拟合效果较好,东山、芗城和龙海 3 个站拟

合效果较差。若取显著性水平 $\alpha=0.05$,在 Excel 2000 中用函数 CHIINV(0.05,1)求得 $\chi^2_{0.05}(1)=3.84$,拟合较好的 7 个站的 χ^2 均小于 3.84,可以认为其年最大日雨量总体服从 P- III分布。拟合较差的 3 个站可通过 c_v 和 c_s 参数调整以取得更好的拟合效果。

表 2　漳州市 10 个站的 χ^2 值

台站	χ^2	台站	χ^2
芗城	6.76	平和	0.44*
龙海	9.25	南靖	2.86*
东山	4.62	华安	1.00*
诏安	1.55*	长泰	3.12*
云霄	2.62*	漳浦	2.60*

注:* 为通过 $\alpha=0.05$ 显著水平检验。

4　小结与讨论

(1) P- III分布具有较好的模拟能力,可用来拟合年最大日雨量的分布,进而求得一定重现期的值。在漳州市 10 个站的拟合中,有 7 个站效果较好,能通过显著性水平 $\alpha=0.05$ 的 χ^2 检验。

(2)应用 Excel 2000 强大的数理统计功能,构造 P- II概率分布函数公式,可避免手工查表求离均系数的繁琐,实现软件的快捷计算。

(3) 对于概率分布,样本数越大则拟合的概率分布越接近总体分布。根据文献和经验,P- III分布要求的样本数通常应在 50 以上,至少要大于 40,当样本数较小时,必须采用较严格的检验。本文的样本数为 43,是因为本地观测数据的局限所至,1960 年以前的数据较为凌乱,在气候应用中,通常取 1961 年以后的数据。

(4)对于最大可能降水的计算,目前我国仍以频率计算法为主,水文气象法则主要用于论证。不同地区暴雨频率适合的分布可能不同,可以尝试用对数 P- II分布、对数正态分布、威布尔或耿贝尔极值分布来拟合。

(5)本文对 P- III分布的参数采用矩法估计,计算出的年最大日雨量不同重现期的值可在一般设计标准下使用。P- III分布参数估计的研究很多,对于设计要求较高的情况,今后应探讨应用混合权函数或线性矩等方法来估计参数,以提高精度。

参考文献

1　魏生生,郭化文,陈建昌.国内外求算可能最大降雨量研究的综述.气象科技,1998,27(1):16 - 21
2　陈建昌,郭化文,魏生生,等.用 Jenkinson 法推算山东年最大日雨量重现期值的初探.应用气象学报,1995,6(4):486 - 491
3　翟裕宗.用 Gumbel 分布计算三江平原地区年最大日雨量重现期值的初步结果.黑龙江八一农垦大学学报,1983,2:83 - 92
4　马开玉.气候统计原理与方法.北京:气象出版社,1993.69 - 76
5　高绍凤,陈万隆,朱超群,等. 应用气候学.北京:气象出版社,2001
6　孙志刚,杨聪. Excel 在经济与数理统计中的应用.北京:中国电力出版社,2004

Estimation of Annual Maximum Diurnal Precipitation for Reappearance Periods with Pearson- III Distribution

Lin Liangwei　Wang Liping
(Zhangzhou Meteorological Bureau, Fujian Province, Zhangzhou 363000)

Abstract: The method for estimating the annual maximum diurnal precipitation in various reappearance periods with Pearson- III distribution is described. The maximum diurnal precipitation for reappearance periods was calculated by using the maximum diurnal precipitation data from 10 stations in Zhangzhou from 1961 to 2003. The calculation was performed on Microsoft's Excel 2000. The results show that the fitting results are satisfactory in 7 out of 10 weather stations, and those for the other three stations are poor. The calculation precision of Pearson- III distribution is discussed.

Key words: Pearson- III distribution, reappearance period, maximum diurnal precipitation, Excel